ELFT-EFS Public Challenge

Questions and answers to date

Last updated 26 June 2009

Q1: In the overview you are mentioning that the benchmark will review "the comparative accuracy of image-only searches, image+minutiae searches, and image+extended feature searches." We fully agree on this approach. Yet the different latent datasets do not seem to correspond to such tests. For instance, does the L2 dataset correspond to image+minutiae, or is it minutiae alone? [Q1a] Our recommendation is image+minutiae, which is state of the art today. What is the purpose of L4 (EFS without the image)? [Q1b]

A1a: The wording was changed to make the latent subtest more explicit: L2 dataset refers to "image with EFTS-LFFS features (fields 9.014-9.023)"; the additional L5 refers to "EFTS-LFFS features alone"

A1b: There are three justifications for the latent subtests without images in the public challenge. Including these subtests in the public challenge does not necessarily indicate that the full ELFT-EFS evaluations should include these subtests: the public challenge results may be used in these cases to indicate whether further evaluations are necessary.

- At the workshop, the participants on the vendor panels indicated that image+feature searches are more accurate than feature searches. This is a commonly accepted assumption, but there is not publicly available evaluation data to support the assumption.
- In some operational cases, including the image in transactions is problematic due to bandwidth. Users in these cases need to understand the accuracy impact of these restrictions.
- There is some interest in whether EFS may be used as the basis for a standard fingerprint template (without the image), and what the impact on accuracy may be in that case.

Q2: Will the second evaluation (SDK) be on the same grounds as ELFT07 Phase II, regarding coding/matching speed and SDK interface?

A2: The ELFT-EFS evaluation will be based on lessons learned from ELFT07 Phase II, and both throughput speed and SDK interface are being reviewed. We expect that ELFT-EFS speed requirements will be faster than ELFT07 Phase II.

Q3: Do you have any idea for test conditions (especially processing time restriction for latent enrollment, for exemplar enrollment and for matching)?

A3: (see A2) We are soliciting input from the participants in the Public Challenge for these time limits:

- Exemplar enrollment (set of 10 rolls)
- Exemplar enrollment (set of unsegmented slaps)
- Latent matching

Q4: Can you clarify the required format of results (score files)?

A4: See the ELFT-EFS Public Challenge Test Plan.

Q5: We would appreciate you including in the document a disclaimer strictly limiting the use of the SDK to ELFT-EFS testing. That is something we had discussed with NIST in previous NIST tests and it is very important to us.

A5: We will, and are discussing specific wording now.

Q6: Can you just download the test sets, perform your own tests and send no results at all (without being notated in any report)?

A6: If you intend to participate in the public challenge, we will give you an ID/Password to download the public challenge data. If you do so we request that you return results from some or all of the public challenge subtests. The public challenge data will be made

public after the completion of the ELFT-EFS Public Challenge. If you want only want the data for internal testing, please wait until then. If you intend to participate but for some reason are unable to return results, you will not be penalized: the participants (including participants who do not return results) will remain anonymous: they will not be listed by name, nor have specific results attributed to them.

Q7: I would like to confirm test condition regarding L2 and L3: - L2 - image with EFTS-LFFS features (fields 9.014-9.023) - L3 - image with EFS features (fields 9.300-9.373) I assume that we can use features (EFTS-LFFS or EFS) when we encode latent from image. For example, we can use manually-encoded direction data (if available) to extract ridge from image. Am I correct?

A7: Yes, both the image and the manually-encoded features may be used in this case. For example, automatic feature extraction from the latent image could be performed (as in your example) using information from the provided manually-encoded features, or not. Note that it has been asserted by several AFIS vendors that their operational systems receive both the image and the manually-encoded features (and it is assumed that the image is somehow utilized in addition the provided features). In contrast, the FBI IAFIS system is thought to ignore the image when image+minutiae are provided, so for systems such as IAFIS which operate in either image-only or feature-only search modes, L2 or L5 would likely produce the same results.

Q8: I understand that this test condition does not support some operational procedures such as: 1) To launch latent matching using image only 2) To launch latent matching with manually encoded data when 1) did not retune Positive Hit 3) To fuse scores of 1) and 2) and to reproduce candidates However, this accuracy can be simulated from the results (scores) of L1 and L4 (or L5). Therefore, additional test is not necessary. [L1: image only L4: EFS features alone L5: EFTS-LFFS features alone]

A8: Generation of performance results based on fusion (conditional or otherwise) of candidate lists is an option that we may pursue.

Q9: Released (already provided) latent data do contain some of the EFS features (fields 9.300-9.373). However, they do not seem to contain any of EFTS-LFFS features (fields 9.014-9.023). Please let me know your intention:

a) Do you intend us (participants) to use latent data on SD27? [i.e. the "ideal" feature markup that was based on both the latent and exemplar] I do not think it is good idea because minutia positions are different, i.e. it is difficult to compare advantage of EFS (additional) features against EFTS-LFFS features.

b) Do you intend us (participants) to use only minutia data (position, direction) on EFS features (fields 9.300-9.373). I think it is better idea than method a). However, I am afraid if it would create unnecessary confusion. For example, how can we determine file name of candidate (score) files?

c) Do you intend release separate latent data files which have EFTS-LFFS features (fields 9.014-9.023) and which are actually copy of the part of EFS features (fields 9.300-9.373). I would like to recommend this method.

A9: EFS is a deliberate superset of the IAFIS/EFTS/LFFS fields. The final version of the latent data will have IAFIS/EFTS/LFFS features that were automatically extracted from the EFS features. [i.e. C is correct.]

Q10: We are having problems extracting the following WSQ images:.

EFS_PC_500ppi_MatedExemplarsREVISED/E078A.AN2 record 03

EFS_PC_500ppi_MatedExemplarsREVISED/E078C.AN2 record 06

EFS_PC_500ppi_MatedExemplarsREVISED/E078C.AN2 record 08

EFS_PC_500ppi_MatedExemplarsREVISED/E078D.AN2 record 02

EFS_PC_500ppi_MatedExemplarsREVISED/E078D.AN2 record 04

EFS_PC_500ppi_MatedExemplarsREVISED/E078D.AN2 record 05

EFS_PC_500ppi_MatedExemplarsREVISED/E078E.AN2 record 05

EFS_PC_500ppi_MatedExemplarsREVISED/E078F.AN2 record 01

EFS_PC_500ppi_MatedExemplarsREVISED/E078F.AN2 record 02

EFS_PC_500ppi_MatedExemplarsREVISED/E078F.AN2 record 03

EFS_PC_500ppi_MatedExemplarsREVISED/E078F.AN2 record 09

EFS_PC_500ppi_MatedExemplarsREVISED/E078H.AN2 record 06

EFS_PC_500ppi_MatedExemplarsREVISED/E078I.AN2 record 01

EFS_PC_500ppi_MatedExemplarsREVISED/E078I.AN2 record 04

EFS_PC_500ppi_MatedExemplarsREVISED/E078I.AN2 record 06

EFS_PC_500ppi_MatedExemplarsREVISED/E078I.AN2 record 09

Provided as part of ULW tools 'ANSINISTViewer.exe' also crashes extracting these images. Could you check mentioned files and confirm if there is an error in WSQ packaging or the error is on our side ?

A10: Some participants have had difficulty in opening a few of the WSQ files (partial list above). The problem turned out to be nonstandard use of comments within the WSQ data, which causes problems for some but not all WSQ decoders.

There should not be a reason to download files unless you have a problem reading the affected files.

In case you are having problems, we have removed all WSQ comments from all files in EFS_PC_500ppi_MatedExemplars. The zipfile:

EFS_PC_500ppi_MatedExemplars20090527.zip
replaces the zipfile
EFS_PC_500ppi_MatedExemplarsREVISED.zip.

Again, there should be no reason to download the data again unless this problem is affecting you.

Q11: In the Test Plan you are only asking participants to report their timings, without giving any guidelines on what a reasonable coding/matching speed should be. Participants might choose very different timings which could hinder the final analysis. As you know the matching accuracy can me much improved given sufficient processing time. Do you plan on releasing speed requirements before phase 1?

A11: Our intent is to use all timing results obtained from the challenge problem to inform the limits placed on future evaluations. Based on feedback from ELFT Phase II, our current assumption for subsequent evaluations (which is subject to change based on these results, feedback, etc.) is that the latent encode/search time limits will be approximately 5x faster than ELFT Phase II.

Q12: What is the purpose of having multiple exemplars for some subjects with respect to the 500ppi mated exemplars provided? How should these be treated with respect to the candidate lists returned?

A12: The purpose is not to evaluate "linked" multiple exemplars (i.e. combined exemplar sets for a given subject), but rather to assess the range of performance obtained in the best, median, and worst case across all exemplars are present. Thus, all exemplars should be searched and the returned results (candidates) for each shall be independent. To be specific: this challenge problem is a case in which we can explore a variety of cases in which a given subject has five, ten, or more exemplars, and compare the results. To that end, for a given search, we intend to show what the results would be if we ignore all mates but a) the highest-scoring, b) the median, and c) the lowest scoring. This can give us some insight (limited, of course by the challenge test size) into the effect of variations between exemplars.

Q13: How should galleries for the searches of E1, E2, E3, and E4 be constructed and searched?

A13: Before a specific exemplar subtest can be executed, an appropriate gallery must be constructed containing both mated (foreground) and non-mated (background) exemplars appropriate to the subtest. As stated in the Public Challenge Problem Test Plan, the following are the Exemplar subtests:

- E1 - 1000ppi rolled exemplars
- E2 - 500ppi rolled exemplars
- E3 - 1000ppi plain exemplars (unsegmented slaps)
- E4 - 500ppi plain exemplars (unsegmented slaps)

Therefore, these are the galleries to be used:

- E1 - 1000ppi rolled exemplars
  - 204 mated exemplar sets from EFS_PC_1000ppi_MatedExemplars
  - 214 non-mated exemplar sets from EFS_PC_1000ppi_Nonmate_Exemplars
  - Only the Rolled prints (pos 1-10) are used from each file
- E2 - 500ppi rolled exemplars
  - 643 mated exemplar sets from EFS_PC_500ppi_MatedExemplars
  - 214 non-mated exemplar sets from EFS_PC_500ppi_Nonmate_Exemplars
  - Only the Rolled prints (pos 1-10) are used from each file
- E3 - 1000ppi plain exemplars (unsegmented slaps)
  - 204 mated exemplar sets from EFS_PC_1000ppi_MatedExemplars
  - 214 non-mated exemplar sets from EFS_PC_1000ppi_Nonmate_Exemplars
  - Only the Plain/Slap prints (pos 11-14) are used from each file
- E4 - 500ppi plain exemplars (unsegmented slaps)
  - 643 mated exemplar sets from EFS_PC_500ppi_MatedExemplars
  - 214 non-mated exemplar sets from EFS_PC_500ppi_Nonmate_Exemplars
  - Only the Plain/Slap prints (pos 11-14) are used from each file

The ordering of all exemplars within a gallery should be randomized with the matching software unaware of the mated exemplar positions.

Q14: It appears there are wrong IDs for some latents/exemplars:

Latent 240.01 Hit on Exemplar 215.01

Latent 163.07 Hit on Exemplar 181.07

Latent 181.08 DOES NOT Hit on Exemplar 181.08

A14: The following corrections will be made:

- Latent # 240 belongs to Subject # 215 (there is no subject 240)
- E181 (Subject # 181) was incorrectly labeled – it should be E163. (both 500ppi and 1000ppi images). Latent 181 has no corresponding exemplar.

We will update the crossreference spreadsheet.

Q15: The exemplar L187.an2 is poorly segmented so that the rolled image for finger #2 has problems matching the latent 187.

A15: You are correct. L187.an2 (both 500 and 1000ppi) is poorly segmented -- we received the data this way, and unfortunately this represents a real-world occurrence. Note that the second 500ppi exemplar set (L187A.an2) does not have this problem.

Q16: There is no exemplar for subject 253.

A16: There is no 1000ppi exemplar for subject 253. Note there is an exemplar (E253A.an2) at 500ppi.

Q17: The following exemplars are missing from the datasets (column 2 of "EFS-PC_Xref.txt") for both resolutions.:

Latent_Image_Number	Subject
L038	38
L059	59
L081	81
L085	85
L101	101
L117	117
L118	118
L119	118
L127	101
~~***L163***~~	~~***163***~~ (see Q14)
~~***L240***~~	~~***240***~~ (see Q14)
L251	251
L254	254

A17: Yes, this is consistent with section 2 of "ELFT-EFS-PC_Readme_2009-05-20.pdf". There are 12 latents corresponding to 10 unique subjects without an exemplar mate available at either resolution (500 or 1000 ppi):

- L038
- L059
- L081
- L085
- L101
- L117
- L118
- L119
- L127
- L181 (not L163 – see Q14)
- L251
- L254

Note: As stated in Q16, one latent (L253) has one 500ppi exemplar mate (E253A), but no 1000ppi exemplars.

Therefore, when searched against the 1000ppi gallery there are 13 latent searches which cannot be mated, and 12 searches which cannot be mated against the 500ppi gallery.

Q18: The mate-images at 500 PPI appear to be slightly blocky. Was the compression/decompression done correctly? (For example, most of the images in \EFS_PC_500ppi_Exemplars_full3\). However, the 1000 PPI images appear to be much better.

A18: The 1000ppi images were scanned and compressed by NIST. With the 500ppi images, there are two sources:

- The 500ppi images without a suffix (e.g. E002.an2, E003.an2) are the same images as the 1000ppi images (reduced by pixel sampling, not pixel averaging), and were compressed by NIST
- The 500ppi images ***with*** a suffix (e.g. E003A.an2, E005L.an2) were obtained from various sources, and the compression methods used were outside of NIST's control. For good or bad, these are representative of reality. Note that recompression is not possible without further increasing artifacts.

Q19: About the L1- image (only) test set... Some of the latents have multiple images in them — so, processing and matching the whole image may not be very meaningful (e.g., ID's 56, 57, 58, etc.). If one were to use the masks from the other sets — which one to use (the Juried one perhaps?)

A19: As several people discussed in the NIST latent workshop, there are multiple definitions for image-only searching, and multiple scenarios for use. Among these:

A. Automated searching of latent images without human preprocessing (i.e. image only, no region of interest)
B. Automated searching of latent images with minimal human preprocessing (i.e. image only, region of interest and orientation marked)

There are operational needs for both processes. For the public challenge dataset, we decided not to further increase complexity by adding another latent subtest: the L1 test is just for the image (A), not ancillary information (B). For the actual evaluation, we plan to run subtests for both image only and image+region of interest.

Q20: Regarding the Answer to Question 19 of the last FAQ of June 3rd 2009. It seems to me that for the matching of the latent images in set L1, no "human interaction" with the images is allowed whatsoever. By "Human Interaction" I mean manually editing the minutiae of a given image, including the manual definition of a Region of Interest. Is my understanding correct?

A20: Yes, you are correct. For L1, the intent is image-only searches without any manually created additional information (e.g. region of interest). The idea with L1 is to measure how well the software is able to perform image-only matching without any human assistance.

Q21: Finger No. for plain exemplars (unsegmented slaps) (E3 and E4) in the candidate lists. I am assuming that we shall record 1-10 (Finger Pos.) based on segmented data for the Finger No. instead of 11-14. Am I correct?

A21: Correct.

Q22: For the submission method of the public challenge results, I am assuming that I can send our results over e-mail after encryption. Am I correct?

A22: Please put all of the candidate lists for each subtest in a separate zipfile, named <vendor>_L#xE#.zip, emailed to latent-efs@nist.gov using zip encryption and the password used for https://tsf.noblis.org/download/ELFT-EFS-PC/.

- For example, if Vendor FOO is providing results for all tests, then 20 zipfiles will be provided: FOO_L1xE1.zip, FOO_L1xE2.zip, […], FOO_L5xE4.zip.
- The zipfiles can be encrypted using any of these common encryption methods (listed in descending encryption strength): 256-bit AES, 128-bit AES, Zip 2.0.

Q23: A question came up today that relates to the sequestered test and your goals for the testing: There are several different sub-tests to this EFS test, and some problems may be better solved with different approaches. For instance the minutiae extractor used for the gallery exemplars may differ for image-only lights out (L1) tests compared to the other tests. This is fine when we are just submitting results since we could enroll different gallery templates for each sub-test. However, for the sequestered test, this would require you to perform different enrollments for the different scenarios tested. Should we constrain our approach to a single enrollment, or in the sequestered test will we have the option to enroll different templates for the different types of tests?

A23: In the sequestered test, there will only be a single opportunity for enrollment. However, the "template" that results from that enrollment may in practice be a superset of different special-purpose templates from your point of view. When you provide feedback and recommendations at the end of the Public Challenge, you may wish to revisit the template size and/or enrollment speed implications of this in your comments.

Q24: Please note that the files containing EFS features (sets L3 and L4) cannot be opened using ULW 5.6. When is the version available?

A24: ULW Beta release 5.6.8 may be downloaded from the following site: https://tsf.noblis.org/download/ELFT-EFS-PC/ULW_568.zip [Access requires the UserID and password (case sensitive) that we sent to you previously.] This version of ULW is intended for EFS testing and should not be used for operational work.

Q25: Can you please clarify how you intend to calculate rank statistics when there exists more than 1 exemplar for a latent? There are generally 2 ways: take only the highest exemplar rank or use all exemplars from the same finger separately (ignoring each other) in calculating the rank statistics. The first method would reflect the 'hit rate' and would allow pooling of the gallery (exemplar) data. The other method would reflect the accuracy if a database de-duplication was needed or would better reflect the effect of removing one of the exemplars.

A25: For rank-based analysis against the 1000ppi galleries (E1/E3), the rank-based analyses will consists of a CMC over ranks 1-100 for all searches having a mate in the gallery.

For rank-based analysis against the 500ppi galleries (E2/E4), many subjects have multiple exemplars per subject, and therefore CMC/rank-based analysis must use a specialized approach. For the 500ppi exemplars, there will be 4 rank-based analyses:

Baseline Identification Rate (Comparable to 1000ppi Results)

This models a scenario where the gallery retains a single encounter, specifically that corresponding to the 1000 ppi gallery. Results will be scored as if only the one corresponding impression were retained for each subject_finger. All additional exemplars per-subject beyond those used for the 1000ppi gallery will be discarded from the candidate lists before computing the mate rank. Note that the 500ppi images without a suffix (e.g. E002.an2, E003.an2) are the baseline images – subsampled versions of the 1000ppi images.

Identification Rate with Multiple Encounters: BEST RANK

This models an operational scenario where the gallery retains multiple encounters and matchers return candidate lists of subjects. Results will be scored as if only one impression were retained for each subject_finger. The highest scoring mate and the highest scoring non-mates will be retained with all other candidates discarded from the candidate list before computing the mate rank.

Identification Rate with Multiple Encounters: BEST NFIQ

This models an operational scenario where the gallery retains a single encounter based on highest NFIQ score. Results will be scored as if only one impression were retained for each subject_finger. The "best" impression by NFIQ for each subject_finger (breaking ties arbitrarily) will be retained, with all others discarded from the candidate list before computing the mate rank.

Identification Rate with Multiple Encounters: RANDOM

This models a scenario where the gallery retains a single encounter selected at random. Multiple lists of one impression (selected at random) per subject will be created. Results will be scored as if only one impression were retained for each subject corresponding to the impression in each list. Multiple overlayed CMCs will be generated, one per list, for each subtest. The aggregate of these multiple CMCs will show the variation in accuracy due to different exemplars.

Q26: In the L3 set, print L006G, all three markups have orientation direction = 19, and uncertainty = 16, which I thought was rather remarkable for that latent. (Q26a) I was surprised that humans could specify these numbers so accurately (I assume the examiners sort of guessed the orientation). Also, since the default value of u = 15, it is remarkable that the value of u = 16 was even specified. (Q26b)

A26a: During the initial markup for this dataset, the examiners did not mark the orientation. When the images were juried, the orientation was marked and that orientation was pushed to the other markups. Therefore for many of the images in the public challenge dataset, all three markups have the same orientation.

A26b: The orientation and uncertainty were determined graphically by mouse dragging – the examiners dragged a line to indicate upright, and dragged a mirrored pair of lines to indicate uncertainty. The examiners did not type in the numbers.

Q27: Regarding field 9.308 (RQM): 6 valid values (quality codes) are defined (0 through 5) but only 2 bits are allocated, which are not enough for representation of 6 values. How would this be resolved?

A27: It appears that you are referring to the EFS draft 0.2 or earlier, which had the issue you mention for one format. In EFS 0.3 and 0.4, field 9.308 can use the formats specified infield 9.309, Uncompressed or Run-Length Encoded. The data for the ELFT-EFS evaluation will use the Uncompressed format for simplicity, which uses one character

per grid cell. The current specification can be obtained from http://fingerprint.nist.gov/standard/cdeffs/

Q28: It is stated in "Public Challenge Problem Test Plan' (04/18/09) that "… participants will remain anonymous in the results". Can you please confirm this?

A28: The results for the Public Challenge will be anonymous: they will not indicate the participants' names. Participants are not prohibited from disclosing their results from the Public Challenge.

Q29: In the Public Challenge dataset, a few of the cells for the ridge flow (9.310) seem to be off. An example of marked extended data with 'off' value: L3\L002G_JURIED_NP.lffs, Row 54 , 25th block orientation is 0x5A=90, and 26th block orientation is 0x16=22. The difference of these two angles is 90-22=68. It is in an area where ridges go smoothly. There is no data for the 24th block while the orientation for the 23th block is 45.

A29: The ridge flow values were extracted from the ridge tracing. The tracing in the cell mentioned in your example has a small downward bend, which resulted in the ridge flow angle you note. The algorithm currently used is based on the current cell direction, without using a larger window, which results in sensitivity to local changes. In discussions among CDEFFS (in 2006), the committee considered and rejected recommending/requiring a specific window size for determining ridge flow. One implication of using a small window size is that the consumer of the ridge flow data may choose to blur the data (effectively applying a larger window after the fact), but has the option of choosing the window size to use.

Q30: Your 06/05 email mentions a 'public challenge report'(again, the participants will remain anonymous in the report) that is to be made public. Is this still the plan? If yes, when would it be available?

A30: Yes. We intend to make draft results available to the participants by the end of July, before we accept the latent matcher SDKs for Evaluation #1.

Q31: Our understanding is that a technology provider can participate in public challenge but opt to not participate in evaluation #1. Is this correct? If so, does the provider have to inform NIST about their withdrawing before a certain deadline to remain anonymous?

A31: You may participate in the Public Challenge and not participate in Evaluation #1. Public challenge anonymity is not related to participation in Evaluation #1.

Q32: Our understanding is that data in fields 9.310 and 9.372 are not marked manually but generated by software. Is this correct? If yes, (1) can you please let us know what software is used to generate the data? and (2) Are there any other fields that are not marked manually?

A32: The skeletonized image (9.372) WAS marked manually. The ridge flow map (9.310) was generated from the skeleton using the ULW libraries — the ridge flow map is the only field that was generated by software rather than direct human markup.